ECON 120: Section Notes

Week 4

David Sungho Park

October 26, 2020

General tips for reading papers

- What are the main findings?
- What is the estimation method? (e.g. RCT, difference-in-differences, etc.)
  - What is the main assumption for the method to be valid?
  - How could it fail?
- What is the identification strategy? In other words, how did the intervention roll out? Randomly or not?
- Why might the estimated effects be biased or unbiased?
  - What kind of selection bias, if any?

Fall '19 midterm - Q3

3. You are evaluating a program to provide scholarships to encourage kids to stay enrolled until the end of primary school (up to grade 8). Imagine that there are three types of kids in these schools. Call them Types A, B, and C. 50% of kids are Type A, 25% are Type B, and 25% are Type C. Type A kids are rich and will finish school with or without a scholarship. Type B kids are very poor and won't finish school with or without a scholarship. Type C kids are somewhere in the middle, and will finish school if they get a scholarship but will drop out otherwise. You are evaluating an experiment in which kids in treatment schools get scholarships. Kids in control schools get nothing.

(A) (50%) : will finish school regardless of scholarship

(B) (25%) : won't finish w/ or w/o scholarship

(C) (25%) : will finish only if scholarship

$\frac{2.5 \text{ point}}{0.25} = 10 \text{ points}$

Fall '19 midterm - Q3.(a)

(a) Write out the regression you'd run to test whether the program increased school completion rates. Based on the description above, what will your coefficient estimates be?

$$Completion_i = \beta_0 + \beta_1 Scholarship_i + \varepsilon_i$$

$\beta_0 = 0.50$

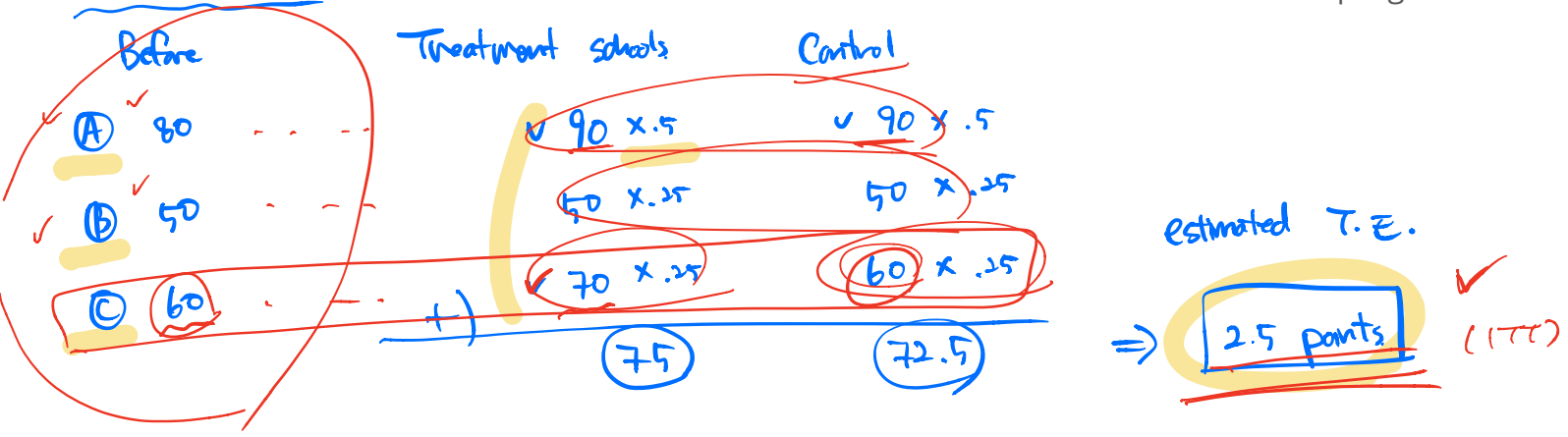
$\beta_1 = 0.25$

β_0 : avg. completion rate of control schools

$\beta_0 + \beta_1$: avg. " " of treatment schools

Fall '19 midterm - Q3.(b)

(b) You are interested in testing whether the program increased test scores. Before the program started, you collected test scores for all the kids. The scores range from 0-100. On average, Type A kids got an 80, Type B kids got a 50, and Type C kids got a 60. Let's assume that any kid who finishes primary school increases their score by 10 points, while any kid who drops out gets the same score that they did before the program started. In the experiment, you conduct a test score that is administered to every single student from the experiment. What is the average post-test score in the treatment schools? What about the control schools? What is the estimated effect of the program?

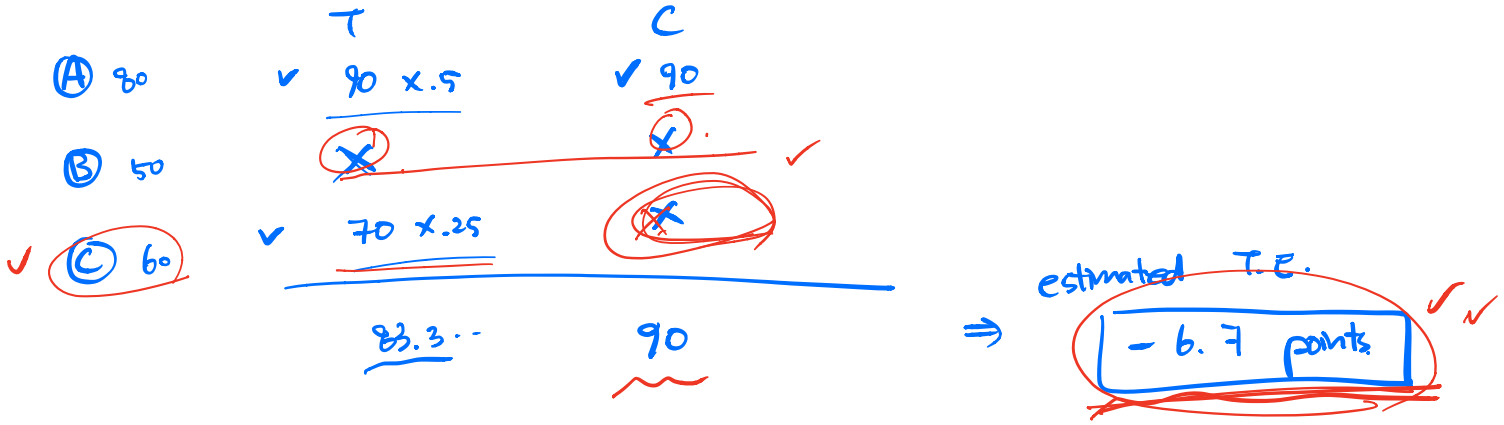


Fall ‘19 midterm - Q3.(c)

(c) Does this estimate give you the return to attending primary school? If not, is this estimate too big or too small? Explain what’s happened.

Fall '19 midterm - Q3.(d) & (e)

(d) Now imagine that the tests can only be administered in the schools. If the kid is still enrolled, she takes the test; if she's dropped out, she can't be found and is never tested. What would your estimated program effect be? Explain your result.



(e) Summarize what you have found in this problem.

Fall ‘19 midterm - Q4

4. As we have talked about in class, school enrollment in developing countries has increased dramatically over time; however, many kids who are enrolled in school score well below grade-level on tests. While these kids are in school, they may not be learning that much. One reason for this may be that classrooms are big and students have heterogeneous preparation levels, so the teacher may find it hard to meet the needs of every child. For example, some kids may be doing advanced math while others are struggling with much more basic concepts.

In this case, students might benefit from more specialized learning opportunities. While a teacher doesn’t have time to create an individual lesson plan for every student, a computer could. Researchers tried out an evaluation in which children in India went to an after-school tutoring center. At the center, each child had access to his or her own computer. Every day, the student would go through lessons on the computer, with the oversight of a teacher, and then would receive person-to-person tutoring afterwards. The computer software was designed to “Teach at the Right Level.” That is, for a student that was struggling, the lessons would be more basic and would be about developing basic skills; for a stronger student the lessons might be more complicated.

Researchers randomized access to this program. A treatment group received the **computer- assisted learning program**, while a control group did not. This question is about evaluating the effect of the program.

Fall '19 midterm - Q4

(a) (5 points) Table 1 shows a table of means for treatment and control groups at baseline, before the program started. What should this show? What does this table show? What does this mean for the

TABLE 1—SAMPLE DESCRIPTIVES AND BALANCE ON OBSERVABLES

	Mean treatment	Mean control	Difference	Standard error	Observations (treatment)	Observations (control)
Panel A. All students in the baseline sample						
Demographic characteristics						
Female	0.76	0.76	0.004	0.034	314	305
Age (years)	12.67	12.41	−0.267	0.143	230	231
SES index	−0.03	0.04	−0.070	0.137	314	305
Grade in school						
Grade 4	0.01	0.01	−0.003	0.007	305	299
Grade 5	0.01	0.02	−0.007	0.010	305	299
Grade 6	0.27	0.30	−0.035	0.037	305	299
Grade 7	0.26	0.26	0.005	0.036	305	299
Grade 8	0.30	0.28	0.017	0.037	305	299
Grade 9	0.15	0.13	0.024	0.028	305	299
Baseline test scores						
Math	−0.01	0.01	−0.016	0.081	313	304
Hindi	0.05	−0.05	0.096	0.080	312	305
Present at endline	0.85	0.90	−0.048	0.027	314	305
Panel B. Only students present in endline						
Demographic characteristics						
Female	0.77	0.76	0.013	0.036	266	273
Age (years)	12.61	12.37	0.243	0.156	196	203
SES index	−0.17	0.03	−0.193	0.142	266	273
Grade in school						
Grade 4	0.01	0.01	−0.003	0.008	258	269
Grade 5	0.01	0.02	−0.011	0.011	258	269
Grade 6	0.28	0.30	−0.022	0.040	258	269
Grade 7	0.26	0.26	−0.001	0.038	258	269
Grade 8	0.30	0.28	0.020	0.040	258	269
Grade 9	0.14	0.12	0.017	0.029	258	269
Baseline test scores						
Math	−0.03	−0.00	−0.031	0.086	265	272
Hindi	0.05	−0.07	0.124	0.084	266	273

Notes: Treatment and control groups refer to whether students were randomly assigned to receive an offer of a Mindspark voucher. Variables used in this table are from the baseline data collection in September 2015. The data collection consisted of two parts: (i) a self-administered student survey, from which demographic characteristics are taken and (ii) assessment of skills in math and Hindi, administered using pen-and-paper tests. Tests were designed to cover wide ranges of achievement and to be linked across grades, as well as between baseline and endline assessments, using common items. Scores are scaled here using Item Response theory models and standardized to have a mean of zero and standard deviation of one in the baseline. *SES index* refers to a wealth index generated using the first component from a Principal Components Analysis consisting of indicators for ownership of various consumer durables and services in the household.

$t = \frac{0.004}{0.034} < 1$

Fall ‘19 midterm - Q4

(b) (hard question - 10 points) Figure 1 shows us a graph where the y-axis is the assessed level of achievement of a student, i.e. their actual level of knowledge. We would hope that a child’s level would be at their grade level – i.e. that a 6th grade student has skills which meet expectations for a 6th grader. However, kids that have fallen behind would be below grade-level – a 6th grader may only be able to read at a 4th grade level for example, and so her assessed level would be 4th grade even though she’s in 6th grade. In the graph, the x-axis is the student’s actual grade level, i.e. the grade that they are enrolled in school.

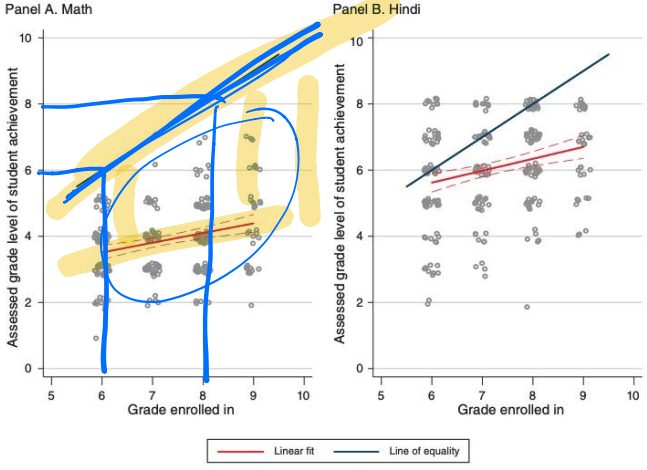


FIGURE 1. ASSESSED LEVELS OF STUDENT ACHIEVEMENT VERSUS CURRENT GRADE ENROLLED IN SCHOOL

i. This graph has a blue line which is the 45 degree line from the origin. What does this line represent?

ii. Where would we expect the points in this graph to fall relative to the 45 degree line?

iii. Interpret the results in this figure.

Fall ‘19 midterm - Q4

(c) (5 points) Table 2 shows us treatment effects from the program. This is a simple regression where the dependent variable is the student’s score on a test after the program was over. The main independent variable is an indicator variable equal to 1 if the student is in the treatment group and 0 otherwise. What does this regression tell us? (note: the dependent variable “standardized IRT score” is just a name for the standardized test score – this variable has mean 0 and standard deviation 1).

TABLE 2—INTENT-TO-TREAT (ITT) EFFECTS IN A REGRESSION FRAMEWORK

	Standardized IRT scores (endline)			
	Math (1)	Hindi (2)	Math (3)	Hindi (4)
Treatment	0.37	0.23	0.37	0.24
	(0.064)	(0.062)	(0.064)	(0.071)
Baseline score	0.58	0.71	0.57	0.68
	(0.042)	(0.040)	(0.051)	(0.033)
Constant	0.33	0.17	0.32	0.17
	(0.044)	(0.044)	(0.031)	(0.035)
Strata fixed effects	Yes	Yes	No	No
Observations	535	537	535	537
R^2	0.403	0.493	0.397	0.473

Notes: Robust standard errors in parentheses. *Treatment* is a dummy variable indicating a randomly assigned offer of a Mindspark voucher. Tests in both math and Hindi were designed to cover wide ranges of achievement and to be linked across grades, as well as between baseline and endline assessments, using common items. Scores are scaled here using Item Response theory models and standardized to have a mean of zero and standard deviation of one in the baseline.

Fall ‘19 midterm - Q4

(d) (5 points) Table 3 shows us how children did on the test on specific items, breaking it down into arithmetic, fractions, etc. What does this show?

TABLE 3—TREATMENT EFFECT BY SPECIFIC COMPETENCE ASSESSED

	Proportion of questions answered correctly						
	Arithmetic computation	Word problems—computation	Data interpretation	Fractions and decimals	Geometry and measurement	Numbers	Pattern recognition
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Mathematics							
Treatment	0.078	0.072	0.042	0.071	0.15	0.15	0.11
	(0.016)	(0.016)	(0.021)	(0.020)	(0.024)	(0.022)	(0.028)
Baseline math score	0.13	0.11	0.082	0.093	0.052	0.068	0.099
	(0.0080)	(0.010)	(0.015)	(0.012)	(0.014)	(0.012)	(0.016)
Constant	0.66	0.50	0.38	0.33	0.39	0.45	0.36
	(0.0079)	(0.0076)	(0.010)	(0.010)	(0.012)	(0.011)	(0.014)
Observations	537	537	537	537	537	537	537
R^2	0.357	0.229	0.097	0.157	0.097	0.135	0.112

	Sentence completion	Retrieve explicitly stated information	Make straight-forward inferences	Interpret and integrate ideas and information
Panel B. Hindi				
Treatment	0.046	0.045	0.065	0.053
	(0.022)	(0.016)	(0.022)	(0.015)
Baseline Hindi score	0.13	0.14	0.15	0.067
	(0.017)	(0.0075)	(0.011)	(0.013)
Constant	0.72	0.59	0.51	0.31
	(0.011)	(0.0078)	(0.011)	(0.0077)
Observations	539	539	539	539
R^2	0.182	0.380	0.309	0.136

Notes: Robust standard errors in parentheses. The tables show the impact of the treatment on specific competences. The dependent variable in each regression is the proportion of questions related to the competence that a student answered correctly. All test questions were multiple choice items with four choices. Baseline scores are IRT scores in the relevant subject from the baseline assessment. *Treatment* is a dummy variable indicating a randomly assigned offer of a Mindspark voucher. All regressions include randomization strata fixed effects.

Fall '19 midterm - Q4

(e) (hard question - 10 points) Table 4 shows us whether treatment effects differ by characteristics like gender. We will focus only on gender, columns 1-2. The regression that the authors are running is

$$testscore_i = \beta_0 + \beta_1 Treat + \beta_3 Female + \beta_4 Female * Treat + \epsilon_i$$

i. What is the prediction from this regression for a boy in the control group? What about a boy in the treatment group?

ii. What is the prediction from this regression for a girl in the control group? What about a girl in the treatment group?

iii. What do each of the coefficient represent?

iv. How would a researcher test whether girls benefit more from the treatment than boys?

Fall ‘19 midterm - Q4

v. What do the results in Table 4 show?

TABLE 4—HETEROGENEITY IN TREATMENT EFFECT BY GENDER, SOCIOECONOMIC STATUS, AND BASELINE SCORE

Covariates	Standardized IRT scores (endline)					
	Female		SES		Baseline score	
	Math (1)	Hindi (2)	Math (3)	Hindi (4)	Math (5)	Hindi (6)
Treatment	0.47	0.27	0.38	0.26	0.37	0.24
	(0.14)	(0.095)	(0.065)	(0.062)	(0.064)	(0.070)
Covariate	−0.050	0.21	−0.0028	0.099	0.53	0.70
	(0.14)	(0.15)	(0.035)	(0.021)	(0.076)	(0.047)
Interaction	*−0.13*	*−0.046*	*0.023*	*−0.0041*	*0.081*	*−0.047*
	(0.14)	*(0.12)*	*(0.050)*	*(0.041)*	*(0.087)*	*(0.071)*
Observations	535	537	535	537	535	537
R^2	0.399	0.474	0.398	0.494	0.399	0.473

Notes: Robust standard errors in parentheses. *Treatment* is a dummy variable indicating a randomly assigned offer of a Mindspark voucher. The SES index and test scores are defined as in Tables 1 and 2 respectively. All regressions include strata fixed effects and control for baseline subject scores.

Fall '19 midterm - Q4

(f) (5 points) Table 6 shows us treatment effects on items that are below a child's grade level, and on items that are at a child's grade level. Interpret these results.

TABLE 6—TREATMENT EFFECT ON ITEMS LINKED TO GRADE LEVELS

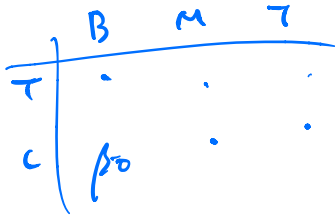
	Proportion of questions answered correctly			
	Math		Hindi	
	At or above grade level (1)	Below grade level (2)	At or above grade level (3)	Below grade level (4)
Treatment	0.0089	0.081	0.063	0.050
	(0.032)	(0.013)	(0.027)	(0.014)
Baseline subject score	0.047	0.099	0.13	0.13
	(0.022)	(0.0069)	(0.016)	(0.0068)
Constant	0.31	0.49	0.45	0.58
	(0.022)	(0.0089)	(0.019)	(0.0100)
Observations	291	511	292	513
R^2	0.029	0.346	0.250	0.399

Notes: Robust standard errors in parentheses. The table shows the impact of the treatment (winning a randomly assigned voucher) on questions below or at/above grade levels for individual students. The dependent variable is the proportion of questions that a student answered correctly. All test questions were multiple choice items with four choices. Our endline assessments, designed to be informative at students' actual levels of achievement, did not include many items at grade 8 level and above. Therefore, students in grades 8 and 9 are not included in regressions on items at/above grade level. Baseline scores are IRT scores in the relevant subject from the baseline assessment. All regressions include randomization strata fixed effects.

Fall '19 midterm - Q4

(g) (hard question - 5 points) Table 5 shows us heterogeneity by tercile of the pre-treatment score. A tercile is a word for splitting up the test into 3 groups – the kids in the bottom third at baseline, the kids in the middle third, and the kids in the top third. Table 5 then runs regressions of this form

$$testscore_i = \beta_0 + \beta_1 Treat + \beta_2 MiddleTercile + \beta_3 TopTercile + \beta_4 MiddleTercile * Treat + \beta_5 TopTercile * Treat + \epsilon_i$$



i. Why is the lowest tercile (and the interaction between the lowest tercile and treatment) omitted here?

ii. Interpret all the regression coefficients.

iii. How would a researcher tell if the treatment worked better for kids at the top tercile than at the bottom?

Fall ‘19 midterm - Q4

iv. What do the results in Table 5 show? Interpret what you find.

TABLE 5—HETEROGENEITY IN TREATMENT EFFECT BY WITHIN-GRADE TERCILES

	Standardized IRT scores (endline)	
	Math	Hindi
	(1)	(2)
Bottom tercile	0.13	−0.072
	(0.098)	(0.10)
Middle tercile	0.30	0.14
	(0.073)	(0.068)
Top tercile	0.53	0.46
	(0.092)	(0.085)
Treatment	0.33	0.41
	(0.12)	(0.12)
Treatment × middle tercile	0.083	−0.30
	(0.16)	(0.16)
Treatment × top tercile	0.068	−0.24
	(0.16)	(0.15)
Baseline test score	0.44	0.58
	(0.066)	(0.062)
Observations	535	537
R^2	0.545	0.545

Notes: Robust standard errors in parentheses. *Treatment* is a dummy variable indicating a randomly assigned offer of a Mindspark voucher. Test scores are scaled as in Table 2.

Problem Set 2 - Question 3